GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 12:46:48 ; Search time 1416 Seconds
(without alignments)
2166.823 Million cell updates/sec

Title: US-09-880-253B-5
Perfect score: 75
Sequence: 1 agacuccagcccuggaccgc.....acacccuccucugagacgcc 75

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		75	100.0	3600	6	BD141660	BD141660 Use of Gl
2		75	100.0	3600	6	BD141661	BD141661 Use of Gl
3		75	100.0	3600	6	BD141662	BD141662 Use of Gl
4		75	100.0	3600	6	BD141663	BD141663 Use of Gl
5		75	100.0	3600	9	HSGLI	X07384 Human mRNA
6		48	64.0	1492	9	AF026306	AF026306 Homo sapi
7	c	48	64.0	166987	2	AC063917	AC063917 Homo sapi
8	c	48	64.0	172136	9	AC022506	AC022506 Homo sapi
9	c	48	64.0	185688	2	AC018805	AC018805 Homo sapi
10		40	53.3	201672	2	AC122965	AC122965 Rattus no
11	c	40	53.3	207478	2	AC108599	AC108599 Rattus no
12		40	53.3	218231	2	AC114111	AC114111 Rattus no
13		39.6	52.8	3587	6	AX676836	AX676836 Sequence
14		39.6	52.8	3587	9	BC013000	BC013000 Homo sapi
15		36.8	49.1	796	10	AF189287	AF189287 Mus muscu
16	c	36.8	49.1	206936	2	AC114678	AC114678 Mus muscu
17	c	34.2	45.6	8513	6	AX251122	AX251122 Sequence
18	c	34.2	45.6	8513	6	AX277897	AX277897 Sequence
19	c	34.2	45.6	8513	6	AX323574	AX323574 Sequence
20	c	34.2	45.6	8513	6	AX344687	AX344687 Sequence
21		31.4	41.9	158811	2	AC013503	AC013503 Homo sapi
22		31.4	41.9	175773	9	AC027128	AC027128 Homo sapi
23	c	31.4	41.9	192973	9	AC023471	AC023471 Homo sapi
24	c	31.4	41.9	222542	9	AC022379	AC022379 Homo sapi
25	c	30	40.0	22970	9	HSL247F6	Z68279 Human DNA s
26	c	30	40.0	174318	9	CNS01RGQ	AL159141 Human chr
27		30	40.0	193371	2	AC098298	AC098298 Rattus no
28		30	40.0	209157	9	CNS01DW4	AL136332 Human chr
29	c	30	40.0	219949	2	AC113819	AC113819 Rattus no
30	c	29.8	39.7	2080	6	AX302031	AX302031 Sequence
31		29.8	39.7	114979	2	AC020798	AC020798 Mus muscu
32		29.8	39.7	139150	9	AC138655	AC138655 Homo sapi
33		29.8	39.7	217304	9	AC009600	AC009600 Homo sapi
34		29.8	39.7	265234	2	AC102562	AC102562 Mus muscu
35	c	29.2	38.9	1620	6	AX440610	AX440610 Sequence
36		29.2	38.9	110000	2	AC091338_0	AC091338 Rattus no
37	c	29.2	38.9	220722	2	AC116257	AC116257 Rattus no
38	c	29.2	38.9	221507	9	HS407F11	AL022329 Human DNA
39		29.2	38.9	233675	2	AC132796	AC132796 Rattus no
40	c	29.2	38.9	234542	2	AC127639	AC127639 Rattus no
41	c	29.2	38.9	236139	2	AC125724	AC125724 Rattus no
42	c	29.2	38.9	237090	2	AC094943	AC094943 Rattus no
43		29.2	38.9	266201	2	AC091336	AC091336 Rattus no
44		29	38.7	385	9	AF005644	AF005644 Pan trogl
45		29	38.7	385	9	AF005645	AF005645 Pan panis

ALIGNMENTS

RESULT 1
BD141660
LOCUS BD141660 3600 bp DNA linear PAT 18-SEP-2002
DEFINITION Use of Gli1 gene.
ACCESSION BD141660
VERSION BD141660.1 GI:23236605
KEYWORDS WO 0211752-A/10.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3600)
AUTHORS Hikichi,Y.
TITLE Use of Gli1 gene
JOURNAL Patent: WO 0211752-A 10 14-FEB-2002;

```
            TAKEDA CHEMICAL INDUSTRIES LTD,YUICHI HIKICHI
COMMENT     OS   Homo sapiens (human)
            PN   WO 0211752-A/10
            PD   14-FEB-2002
            PF   03-AUG-2001 WO 2001JP006688
            PR   04-AUG-2000 JP 00P   242767
            PI   YUICHI HIKICHI
            PC   A61K38/17,A61K48/00,A61K45/00,A61P19/00,A61P19/02,A61P19/10,
            PC   C12Q1/02,
            PC   C12Q1/68,G01N33/15,G01N33/50//C07K14/47,C12N15/12 CC   Use of
            Gli1 gene
            FH   Key             Location/Qualifiers
            FT   source          1. .3600
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .3600
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT      785 a   1161 c    949 g    705 t
ORIGIN

  Query Match             100.0%;  Score 75;  DB 6;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-11;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||:|:||||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 2
BD141661
LOCUS       BD141661                3600 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Use of Gli1 gene.
ACCESSION   BD141661
VERSION     BD141661.1  GI:23236606
KEYWORDS    WO 0211752-A/11.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3600)
  AUTHORS   Hikichi,Y.
  TITLE     Use of Gli1 gene
  JOURNAL   Patent: WO 0211752-A 11 14-FEB-2002;
            TAKEDA CHEMICAL INDUSTRIES LTD,YUICHI HIKICHI
COMMENT     OS   Homo sapiens (human)
            PN   WO 0211752-A/11
            PD   14-FEB-2002
            PF   03-AUG-2001 WO 2001JP006688
            PR   04-AUG-2000 JP 00P   242767
            PI   YUICHI HIKICHI
            PC   A61K38/17,A61K48/00,A61K45/00,A61P19/00,A61P19/02,A61P19/10,
            PC   C12Q1/02,
            PC   C12Q1/68,G01N33/15,G01N33/50//C07K14/47,C12N15/12 CC   Use of
            Gli1 gene
            FH   Key             Location/Qualifiers
            FT   source          1. .3600
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .3600
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT      784 a   1162 c    949 g    705 t
ORIGIN

  Query Match             100.0%;  Score 75;  DB 6;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-11;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||:|:||||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 3
BD141662
LOCUS       BD141662                3600 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Use of Gli1 gene.
ACCESSION   BD141662
VERSION     BD141662.1  GI:23236607
KEYWORDS    WO 0211752-A/12.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3600)
  AUTHORS   Hikichi,Y.
  TITLE     Use of Gli1 gene
  JOURNAL   Patent: WO 0211752-A 12 14-FEB-2002;
            TAKEDA CHEMICAL INDUSTRIES LTD,YUICHI HIKICHI
COMMENT     OS   Homo sapiens (human)
            PN   WO 0211752-A/12
            PD   14-FEB-2002
            PF   03-AUG-2001 WO 2001JP006688
            PR   04-AUG-2000 JP 00P   242767
            PI   YUICHI HIKICHI
            PC   A61K38/17,A61K48/00,A61K45/00,A61P19/00,A61P19/02,A61P19/10,
            PC   C12Q1/02,
            PC   C12Q1/68,G01N33/15,G01N33/50//C07K14/47,C12N15/12 CC   Use of
            Gli1 gene
            FH   Key             Location/Qualifiers
            FT   source          1. .3600
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .3600
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
BASE COUNT      786 a   1161 c    948 g    705 t
ORIGIN

  Query Match             100.0%;  Score 75;  DB 6;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-11;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||:|:||||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 4
BD141663
LOCUS       BD141663                3600 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Use of Gli1 gene.
ACCESSION   BD141663
VERSION     BD141663.1  GI:23236608
KEYWORDS    WO 0211752-A/13.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 3600)
  AUTHORS   Hikichi,Y.
  TITLE     Use of Gli1 gene
  JOURNAL   Patent: WO 0211752-A 13 14-FEB-2002;
            TAKEDA CHEMICAL INDUSTRIES LTD,YUICHI HIKICHI
COMMENT     OS   Homo sapiens (human)
            PN   WO 0211752-A/13
            PD   14-FEB-2002
            PF   03-AUG-2001 WO 2001JP006688
            PR   04-AUG-2000 JP 00P   242767
            PI   YUICHI HIKICHI
            PC   A61K38/17,A61K48/00,A61K45/00,A61P19/00,A61P19/02,A61P19/10,
            PC   C12Q1/02,
            PC   C12Q1/68,G01N33/15,G01N33/50//C07K14/47,C12N15/12 CC   Use of
            Gli1 gene
            FH   Key             Location/Qualifiers
            FT   source          1. .3600
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .3600
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      785 a   1162 c    948 g    705 t
ORIGIN

  Query Match           100.0%;  Score 75;  DB 6;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-11;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||:|||||||:|||||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
                |:||:|:|||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 5
HSGLI
LOCUS       HSGLI                   3600 bp    mRNA    linear   PRI 12-SEP-1993
DEFINITION  Human mRNA for GLI protein.
ACCESSION   X07384
VERSION     X07384.1  GI:31767
KEYWORDS    GLI protein; zinc finger protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3600)
  AUTHORS   Kinzler,K.W., Ruppert,J.M., Bigner,S.H. and Vogelstein,B.
  TITLE     The GLI gene is a member of the Kruppel family of zinc finger
            proteins
  JOURNAL   Nature 332 (6162), 371-374 (1988)
  MEDLINE   88175051
   PUBMED   2832761
REFERENCE   2  (bases 1 to 3600)
  AUTHORS   Kinzler,K.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-1988)
FEATURES             Location/Qualifiers
     source          1. .3600
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="D-259 MG"
                     /clone_lib="lambda phage clones M5E, H31, I2G, J36"
     CDS             79. .3399
                     /note="GLI protein (AA 1-1106)"
                     /codon_start=1
                     /protein_id="CAA30297.1"
                     /db_xref="GI:31768"
                     /db_xref="SWISS-PROT:P08151"
                     /translation="MFNSMTPPPISSYGEPCCLRPLPSQGAPSVGTEGLSGPPFCHQA
                     NLMSGPHSYGPARETNSCTEGPLFSSPRSAVKLTKKRALSISPLSDASLDLQTVIRTS
                     PSSLVAFINSRCTSPGGSYGHLSIGTMSPSLGFPAQMNHQKGPSPSFGVQPCGPHDSA
                     RGGMIPHPQSRGPFPTCQLKSELDMLVGKCREEPLEGDMSSPNSTGIQDPLLGMLDGR
                     EDLEREEKREPESVYETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWGGCS
                     RELRPFKAQYMLVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCE
                     HEGCSKAFSNASDRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHGPDAH
                     VTKRHRGDGPLPRAPSISTVEPKREREGGPIREESRLTVPEGAMKPQPSPGAQSSCSS
                     DHSPAGSAANTDSGVEMTGNAGGSTEDLSSLDEGPCIAGTGLSTLRRLENLRLDQLHQ
                     LRPIGTRGLKLPSLSHTGTTVSRRVGPPVSLERRSSSSSSISSAYTVSRRSSLASPFP
                     PGSPPENGASSLPGLMPAQHYLLRARYASARGGGTSPTAASSLDRIGGLPMPPWRSRA
                     EYPGYNPNAGVTRRASDPAQAADRPAPARVQRFKSLGCVHTPPTVAGGGQNFDPYLPT
                     SVYSPQPPSITENAAMDARGLQEEPEVGTSMVGSGLNPYMDFPPTDTLGYGGPEGAAA
                     EPYGARGPGSLPLGPGPPTNYGPNPCPQQASYPDPTQETWGEFPSHSGLYPGPKALGG
                     TYSQCPRLEHYGQVQVKPEQGCPVGSDSTGLAPCLNAHPSEGPPHPQPLFSHYPQPSP
                     PQYLQSGPYTQPPPDYLPSEPRPCLDFDSPTHSTGQLKAQLVCNYVQSQQELLWEGGG
                     REDAPAQEPSYQSPKFLGGSQVSPSRAKAPVNTYGPGFGPNLPNHKSGSYPTPSPCHE
                     NFVVGANRASHRAAAPPRLLPPLPTCYGPLKVGGTNPSCGHPEVGRLGGGPALYPPPE
                     GQVCNPLDSLDLDNTQLDFVAILDEPQGLSPPPSHDQRGSSGHTPPPSGPPNMAVGNM
                     SVLLRSLPGETEFLNSSA"
     misc_feature    781. .1250
                     /note="zinc finger region (AA 235-393)"
BASE COUNT      785 a   1161 c    949 g    705 t
ORIGIN

  Query Match           100.0%;  Score 75;  DB 9;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-11;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||:|||||||:|||||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
                |:||:|:|||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 6
AF026306
LOCUS       AF026306                1492 bp    DNA     linear   PRI 23-SEP-1999
DEFINITION  Homo sapiens zinc finger transcription factor GLI (GLI) gene,
            5'UTR, partial sequence.
ACCESSION   AF026306
VERSION     AF026306.2  GI:5919240
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1492)
  AUTHORS   Liu,C.Z., Yang,J.T., Yoon,J.W., Villavicencio,E., Pfendler,K.,
            Walterhouse,D. and Iannaccone,P.
  TITLE     Characterization of the promoter region and genomic organization of
            GLI, a member of the Sonic hedgehog-Patched signaling pathway
  JOURNAL   Gene 209 (1-2), 1-11 (1998)
  MEDLINE   98192509
   PUBMED   9524201
REFERENCE   2  (bases 1 to 1492)
  AUTHORS   Liu,C.Z., Yang,J.T., Yoon,J.W., Walterhouse,D. and Iannaccone,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1997) Pediatrics, Northwestern University Medical
            School/Children's Memorial Hospital, 2300 Children's Plaza,
            Chicago, IL 60614, USA
REFERENCE   3  (bases 1 to 1492)
  AUTHORS   Liu,C.Z., Yang,J.T., Yoon,J.W., Walterhouse,D. and Iannaccone,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) Pediatrics, Northwestern University Medical
            School/Children's Memorial Hospital, 2300 Children's Plaza,
            Chicago, IL 60614, USA
  REMARK    Sequence update by submitter
COMMENT     On Sep 23, 1999 this sequence version replaced gi:3004846.
```

```
FEATURES             Location/Qualifiers
     source          1. .1492
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q13-q14.3"
     gene            1. .>1492
                     /gene="GLI"
     promoter        1. .1016
                     /gene="GLI"
     mRNA            1017. .>1067
                     /gene="GLI"
                     /product="zinc finger transcription factor GLI"
     5'UTR           1017. .>1067
                     /gene="GLI"
BASE COUNT      293 a    415 c    471 g    313 t
ORIGIN

  Query Match             64.0%;  Score 48;  DB 9;  Length 1492;
  Best Local Similarity   93.8%;  Pred. No. 0.0013;
  Matches   45;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||||:|||||||:||||||||||:||||||||||||||||||||||
Db       1020 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 1067


RESULT 7
AC063917/c
LOCUS       AC063917              166987 bp    DNA     linear   HTG 21-SEP-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-772E1, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC063917
VERSION     AC063917.22  GI:21431082
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166987)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166987)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (22-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 166987)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 17, 2002 this sequence version replaced gi:20335555.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAUR
                Center clone name: RP11-772E1
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 34% of reads
                Chemistry: Dye-terminator Big Dye: 66% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 158154 bases at least Q40
                Consensus quality: 162225 bases at least Q30
                Consensus quality: 164022 bases at least Q20
                Estimated insert size: 166471; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2496: contig of 2496 bp in length
            *     2497     2596: gap of unknown length
            *     2597     6803: contig of 4207 bp in length
            *     6804     6903: gap of unknown length
            *     6904    10378: contig of 3475 bp in length
            *    10379    10478: gap of unknown length
            *    10479    16020: contig of 5542 bp in length
            *    16021    16120: gap of unknown length
            *    16121    22848: contig of 6728 bp in length
            *    22849    22948: gap of unknown length
            *    22949    31409: contig of 8461 bp in length
            *    31410    31509: gap of unknown length
            *    31510    39101: contig of 7592 bp in length
            *    39102    39201: gap of unknown length
            *    39202    52864: contig of 13663 bp in length
            *    52865    52964: gap of unknown length
            *    52965    67189: contig of 14225 bp in length
            *    67190    67289: gap of unknown length
            *    67290    76119: contig of 8830 bp in length
            *    76120    76219: gap of unknown length
```

```
            *    76220    89425: contig of 13206 bp in length
            *    89426    89525: gap of unknown length
            *    89526   109142: contig of 19617 bp in length
            *   109143   109242: gap of unknown length
            *   109243   130358: contig of 21116 bp in length
            *   130359   130458: gap of unknown length
            *   130459   166987: contig of 36529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166987
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-772E1"
BASE COUNT    43975 a  40017 c  40760 g  40911 t   1324 others
ORIGIN

  Query Match             64.0%;  Score 48;  DB 2;  Length 166987;
  Best Local Similarity   93.8%;  Pred. No. 0.00049;
  Matches   45;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||||:|||||||:||||||||||:||||||||||||||||||||||||
Db     154715 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 154668


RESULT 8
AC022506/c
LOCUS       AC022506               172136 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens 12 BAC RP11-181L23 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC022506
VERSION     AC022506.38  GI:24418014
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172136)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172136)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (05-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 172136)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (11-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 172136)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 172136)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (26-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 172136)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (15-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 29, 2002 this sequence version replaced gi:22779458.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
```

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .172136
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-181L23"
     misc_feature    complement(1. .1999)
                     /note="overlaps bases 6. .2004 of clone AC025165"
                     /function="clone overlap"
     repeat_region   763. .811
                     /rpt_family="G-rich"
     STS             2773. .3082
                     /standard_name="SHGC-148470"
     repeat_region   3093. .3210
                     /rpt_family="G-rich"
     repeat_region   4627. .4649
                     /rpt_family="AT_rich"
     repeat_region   complement(8008. .8311)
                     /rpt_family="AluJo"
     repeat_region   complement(9542. .9847)
                     /rpt_family="AluSx"
     repeat_region   9889. .10004
                     /rpt_family="FLAM_A"
     repeat_region   10023. .11439
                     /rpt_family="L2"
     repeat_region   complement(11572. .11882)
                     /rpt_family="AluSp"
     repeat_region   complement(12363. .12663)
                     /rpt_family="AluSx"
     repeat_region   complement(12988. .13081)
                     /rpt_family="MIR"
     repeat_region   complement(13082. .13398)
                     /rpt_family="AluSx"
     repeat_region   complement(13399. .13516)
                     /rpt_family="MIR"
     repeat_region   14379. .14410
                     /rpt_family="(A)n"
     repeat_region   14874. .15037
                     /rpt_family="L2"
     repeat_region   16935. .16970
                     /rpt_family="GC_rich"
     repeat_region   complement(17292. .17654)
                     /rpt_family="L2"
     repeat_region   17657. .17909
                     /rpt_family="AluSx"
     repeat_region   17949. .18167
                     /rpt_family="AluSg/x"
     repeat_region   18168. .18244
                     /rpt_family="AluSx"
     repeat_region   18245. .19209
                     /rpt_family="LTR13"
     repeat_region   19210. .19443
                     /rpt_family="AluSx"
     repeat_region   complement(19480. .20171)
                     /rpt_family="L2"
     repeat_region   20325. .20574
                     /rpt_family="AluSx"
     STS             21607. .21773
                     /standard_name="RH103508"
     repeat_region   22151. .22171
                     /rpt_family="(A)n"
     repeat_region   23441. .23494
                     /rpt_family="(TG)n"
     STS             23604. .23839
                     /standard_name="D12S1889"
     repeat_region   complement(25085. .25168)
                     /rpt_family="MIR"
     repeat_region   complement(25816. .25988)
                     /rpt_family="FAM"
     STS             28112. .28315
                     /standard_name="25471"
     repeat_region   complement(28275. .28323)
                     /rpt_family="MER66A"
     repeat_region   complement(29010. .29304)
                     /rpt_family="AluJb"
     repeat_region   complement(30472. .30773)
```

```
  Query Match             64.0%;  Score 48;  DB 9;  Length 172136;
  Best Local Similarity   93.8%;  Pred. No. 0.00048;
  Matches   45;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||||:|||||||:||||||||||:||||||||||||||||||||||||
Db     147932 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 147885
```

```
RESULT 9
AC018805/c
LOCUS       AC018805               185688 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 12 clone RP11-564P5, WORKING DRAFT
            SEQUENCE, 29 unordered pieces.
ACCESSION   AC018805
VERSION     AC018805.4  GI:8568931
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185688)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185688)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:6855245.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0564P05
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 87%
            Sequencing vector: plasmid; 13%
            Chemistry: Dye-primer ET; 87% of reads
            Chemistry: Dye-terminator Big Dye; 13% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 172728 bases at least Q40
            Consensus quality: 176882 bases at least Q30
            Consensus quality: 179092 bases at least Q20
            Insert size: 217000; agarose-fp
            Insert size: 182888; sum-of-contigs
            Quality coverage: 3.63 in Q20 bases; agarose-fp
            Quality coverage: 4.26 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1241: contig of 1241 bp in length
            *     1242     1341: gap of unknown length
```

```
*     1342     3280: contig of 1939 bp in length
*     3281     3380: gap of unknown length
*     3381     6298: contig of 2918 bp in length
*     6299     6398: gap of unknown length
*     6399     7915: contig of 1517 bp in length
*     7916     8015: gap of unknown length
*     8016    10375: contig of 2360 bp in length
*    10376    10475: gap of unknown length
*    10476    12815: contig of 2340 bp in length
*    12816    12915: gap of unknown length
*    12916    14982: contig of 2067 bp in length
*    14983    15082: gap of unknown length
*    15083    17219: contig of 2137 bp in length
*    17220    17319: gap of unknown length
*    17320    19673: contig of 2354 bp in length
*    19674    19773: gap of unknown length
*    19774    22632: contig of 2859 bp in length
*    22633    22732: gap of unknown length
*    22733    26670: contig of 3938 bp in length
*    26671    26770: gap of unknown length
*    26771    28734: contig of 1964 bp in length
*    28735    28834: gap of unknown length
*    28835    32050: contig of 3216 bp in length
*    32051    32150: gap of unknown length
*    32151    36565: contig of 4415 bp in length
*    36566    36665: gap of unknown length
*    36666    40956: contig of 4291 bp in length
*    40957    41056: gap of unknown length
*    41057    46020: contig of 4964 bp in length
*    46021    46120: gap of unknown length
*    46121    51188: contig of 5068 bp in length
*    51189    51288: gap of unknown length
*    51289    57149: contig of 5861 bp in length
*    57150    57249: gap of unknown length
*    57250    64232: contig of 6983 bp in length
*    64233    64332: gap of unknown length
*    64333    71020: contig of 6688 bp in length
*    71021    71120: gap of unknown length
*    71121    77194: contig of 6074 bp in length
*    77195    77294: gap of unknown length
*    77295    83886: contig of 6592 bp in length
*    83887    83986: gap of unknown length
*    83987    92361: contig of 8375 bp in length
*    92362    92461: gap of unknown length
*    92462   100887: contig of 8426 bp in length
*   100888   100987: gap of unknown length
*   100988   119696: contig of 18709 bp in length
*   119697   119796: gap of unknown length
*   119797   130942: contig of 11146 bp in length
*   130943   131042: gap of unknown length
*   131043   144137: contig of 13095 bp in length
*   144138   144237: gap of unknown length
*   144238   163904: contig of 19667 bp in length
*   163905   164004: gap of unknown length
*   164005   185688: contig of 21684 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185688
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-564P5"
     misc_feature    1. .1241
                     /note="assembly_name:Contig14"
     misc_feature    1342. .3280
                     /note="assembly_name:Contig16"
     misc_feature    3381. .6298
                     /note="assembly_name:Contig17"
     misc_feature    6399. .7915
                     /note="assembly_name:Contig18"
     misc_feature    8016. .10375
                     /note="assembly_name:Contig19"
     misc_feature    10476. .12815
                     /note="assembly_name:Contig20"
     misc_feature    12916. .14982
                     /note="assembly_name:Contig21"
     misc_feature    15083. .17219
                     /note="assembly_name:Contig22"
     misc_feature    17320. .19673
                     /note="assembly_name:Contig23"
     misc_feature    19774. .22632
                     /note="assembly_name:Contig24"
     misc_feature    22733. .26670
                     /note="assembly_name:Contig25"
     misc_feature    26771. .28734
                     /note="assembly_name:Contig26"
     misc_feature    28835. .32050
                     /note="assembly_name:Contig27"
     misc_feature    32151. .36565
                     /note="assembly_name:Contig28
                     clone_end:T7
                     vector_side:right"
     misc_feature    36666. .40956
                     /note="assembly_name:Contig29
                     clone_end:SP6
                     vector_side:right"
     misc_feature    41057. .46020
                     /note="assembly_name:Contig30"
     misc_feature    46121. .51188
                     /note="assembly_name:Contig31"
     misc_feature    51289. .57149
                     /note="assembly_name:Contig32"
     misc_feature    57250. .64232
                     /note="assembly_name:Contig33"
     misc_feature    64333. .71020
                     /note="assembly_name:Contig34"
     misc_feature    71121. .77194
                     /note="assembly_name:Contig35"
     misc_feature    77295. .83886
                     /note="assembly_name:Contig36"
     misc_feature    83987. .92361
                     /note="assembly_name:Contig37"
     misc_feature    92462. .100887
                     /note="assembly_name:Contig38"
     misc_feature    100988. .119696
                     /note="assembly_name:Contig39"
     misc_feature    119797. .130942
                     /note="assembly_name:Contig40"
     misc_feature    131043. .144137
                     /note="assembly_name:Contig41"
     misc_feature    144238. .163904
                     /note="assembly_name:Contig42"
     misc_feature    164005. .185688
                     /note="assembly_name:Contig43"
BASE COUNT    47302 a  43488 c  44484 g  47607 t   2807 others
ORIGIN

  Query Match             64.0%;  Score 48;  DB 2;  Length 185688;
  Best Local Similarity   93.8%;  Pred. No. 0.00047;
  Matches   45;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||||:|||||||:||||||||||:||||||||||||||||||||||||
Db     174758 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 174711


RESULT 10
AC122965
LOCUS       AC122965               201672 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-349N19, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC122965
VERSION     AC122965.4  GI:25085863
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 201672)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., *Lorensuhewa*,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 201672)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 201672)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 19, 2002 this sequence version replaced gi:23907835.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GYCT
Center clone name: CH230-349N19
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 175103 bases at least Q40
Consensus quality: 177389 bases at least Q30
Consensus quality: 178394 bases at least Q20
Estimated insert size: 180387; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16227: contig of 16227 bp in length
* 16228 16327: gap of unknown length
* 16328 201672: contig of 185345 bp in length.

FEATURES Location/Qualifiers
source 1. .201672
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-349N19"
misc_feature complement(14639. .15474)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ181092"
misc_feature 16328. .17800
/note="wgs_contig"

BASE COUNT 47238 a 43520 c 42628 g 45829 t 22457 others
ORIGIN

Query Match 53.3%; Score 40; DB 2; Length 201672;
Best Local Similarity 83.3%; Pred. No. 0.084;
Matches 40; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||  :|||||||:||||| ||||:|||||||||| |||||| |||||||
Db      89400 AGTTTCCAGCCCTGGACCACGCATCCCGAGCCCTGCGCCCCGACAGAG 89447
```

RESULT 11
AC108599/c
LOCUS AC108599 207478 bp DNA linear HTG 11-OCT-2002
DEFINITION Rattus norvegicus clone CH230-299C4, WORKING DRAFT SEQUENCE.
ACCESSION AC108599
VERSION AC108599.4 GI:23603444
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```
REFERENCE   1  (bases 1 to 207478)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207478)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 207478)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 9, 2002 this sequence version replaced gi:21737669.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GPNU
                Center clone name: CH230-299C4
            ------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191989 bases at least Q40
                Consensus quality: 193858 bases at least Q30
                Consensus quality: 195042 bases at least Q20
                Estimated insert size: 198002; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   207478: contig of 207478 bp in length.
FEATURES            Location/Qualifiers
     source         1. .207478
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-299C4"
     misc_feature   18. .944
                    /note="clone_boundary
                    clone_end:T7
                    site:MboI
                    end_sequence:RXAJR14TJ"
     misc_feature   complement(197039. .197994)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:MboI
                    end_sequence:RXAJR14TV"
     misc_feature   201151. .202226
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   205425. .207478
                    /note="wgs_end_extension
                    clone_end:Sp6"
BASE COUNT    51498 a  42711 c  44220 g  57557 t  11492 others
ORIGIN

  Query Match             53.3%;  Score 40;  DB 2;  Length 207478;
  Best Local Similarity   83.3%;  Pred. No. 0.084;
  Matches   40;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||  :|||||||:|||||  ||||:|||||||||  ||||||  |||||||
Db      49053 AGTTTCCAGCCCTGGACCACGCATCCCGAGCCCTGCGCCCCGACAGAG 49006


RESULT 12
AC114111
LOCUS       AC114111              218231 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-23P10, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC114111
VERSION     AC114111.5  GI:30579709
```

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 218231)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 218231)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 218231)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23268081.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GLMH
Center clone name: CH230-23P10
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185672 bases at least Q40
Consensus quality: 190509 bases at least Q30
Consensus quality: 193980 bases at least Q20
Estimated insert size: 194985; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 121740: contig of 121740 bp in length
* 121741 121840: gap of unknown length
* 121841 195708: contig of 73868 bp in length
* 195709 195808: gap of unknown length
* 195809 214376: contig of 18568 bp in length
* 214377 214476: gap of unknown length
* 214477 216260: contig of 1784 bp in length
* 216261 216360: gap of unknown length
* 216361 218231: contig of 1871 bp in length.

FEATURES Location/Qualifiers
source 1. .218231
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-23P10"
misc_feature 890. .1591
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH275857"
misc_feature complement(17653. .18480)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH275856"
misc_feature 27885. .29132
/note="wgs_end_extension
clone_end:T7"
misc_feature 32824. .35023
/note="wgs_end_extension
clone_end:T7"
misc_feature 36172. .37544
/note="wgs_end_extension
clone_end:T7"
misc_feature 202535. .204207
/note="wgs_end_extension
clone_end:T7"

```
BASE COUNT    49557 a  48073 c  48428 g  50669 t  21504 others
ORIGIN

  Query Match             53.3%;  Score 40;  DB 2;  Length 218231;
  Best Local Similarity   83.3%;  Pred. No. 0.083;
  Matches   40;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||  :||||||:|||||  ||||:|||||||||  ||||||  |||||||
Db       9123 AGTTTCCAGCCCTGGACCACGCATCCCGAGCCCTGCGCCCCGACAGAG 9170


RESULT 13
AX676836
LOCUS       AX676836               3587 bp    DNA     linear   PAT 27-MAR-2003
DEFINITION  Sequence 29 from Patent WO02103028.
ACCESSION   AX676836
VERSION     AX676836.1  GI:29334445
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Baranova,A.V., Yankovsky,N.K., Kozlov,A.P., Lobashev,A.V. and
            Krukovskaya,L.L.
  TITLE     In silico screening for phenotype-associated expressed sequences
  JOURNAL   Patent: WO 02103028-A 29 27-DEC-2002;
            Biomedical Center (RU)
FEATURES             Location/Qualifiers
     source          1. .3587
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      798 a   1143 c    944 g    702 t
ORIGIN

  Query Match             52.8%;  Score 39.6;  DB 6;  Length 3587;
  Best Local Similarity   80.4%;  Pred. No. 0.26;
  Matches   37;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         30 GCCCAGCGCCCAGACAGAGUGUCCCCACACCCUCCUCUGAGACGCC 75
              || |   ||||||||||||:|:||||||||||:||:|:||||||||
Db          2 GCACGAGGCCCAGACAGAGTGTCCCCACACCCTCCTCTGAGACGCC 47


RESULT 14
BC013000
LOCUS       BC013000               3587 bp    mRNA    linear   PRI 22-AUG-2001
DEFINITION  Homo sapiens, Similar to glioma-associated oncogene homolog (zinc
            finger protein), clone MGC:3590 IMAGE:3531657, mRNA, complete cds.
ACCESSION   BC013000
VERSION     BC013000.1  GI:15278120
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3587)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 11 Row: h Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4885278.
FEATURES             Location/Qualifiers
     source          1. .3587
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:3590 IMAGE:3531657"
                     /tissue_type="Muscle, rhabdomyosarcoma"
                     /clone_lib="NIH_MGC_17"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             48. .3368
                     /codon_start=1
                     /product="Similar to glioma-associated oncogene homolog
                     (zinc finger protein)"
                     /protein_id="AAH13000.1"
                     /db_xref="GI:15278121"
                     /translation="MFNSMTPPPISSYGEPCCLRPLPSQGAPSVGTEGLSGPPFCHQA
                     NLMSGPHSYGPARETNSCTEGPLFSSPRSAVKLTKKRALSISPLSDASLDLQTVIRTS
                     PSSLVAFINSRCTSPGGSYGHLSIGTMSPSLGFPAQMNHQKGPSPSFGVQPCGPHDSA
                     RGGMIPHPQSRGPFPTCQLKSELDMLVGKCREEPLEGDMSSPNSTGIQDPLLGMLDGR
                     EDLEREEKREPESVYETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWGGCS
                     RELRPFKAQYMLVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCE
                     HEGCSKAFSNASDRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHGPDAH
                     VTKRHRGDGPLPRAPSISTVEPKREREGGPIREESRLTVPEGAMKPQPSPGAQSSCSS
                     DHSPAGSAANTDSGVEMTGNAGGSTEDLSSLDEGPCIAGTGLSTLRRLENLRLDQLHQ
                     LRPIGTRGLKLPSLSHTGTTVSRRVGPPVSLERRSSSSSSISSAYTVSRRSSLASPFP
                     PGSPPENGASSLPGLMPAQHYLLRARYASARGGGTSPTAASSLDRIGGLPMPPWRSRA
                     EYPGYNPNAGVTRRASDPAQAADRPAPARVQRFKSLGCVHTPPTVAGGGQNFDPYLPT
                     SVYSPQPPSITENAAMDARGLQEEPEVGTSMVGSGLNPYMDFPPTDTLGYGGPEGAAA
                     EPYGARGPGSLPLGPGPPTNYGPNPCPQQASYPDPTQETWGEFPSHSGLYPGPKALGG
                     TYSQCPRLEHYGQVQVKPEQGCPVGSDSTGLAPCLNAHPSEGPPHPQPLFSHYPQPSP
                     PQYLQSGPYTQPPPDYLPSEPRPCLDFDSPTHSTGQLKAQLVCNYVQSQQELLWEGGG
                     REDAPAQEPSYQSPKFLGGSQVSPSRAKAPVNTYGPGFGPNLPNHKSGSYPTPSPCHE
                     NFVVGANRASHRAAAPPRLLPPLPTCYGPLKVGGTNPSCGHPEVGRLGGGPALYPPPE
                     GQVCNPLDSLDLDNTQLDFVAILDEPQGLSPPPSHDQRGSSGHTPPPSGPPNMAVGNM
                     SVLLRSLPGETEFLNSSA"
BASE COUNT      798 a   1143 c    944 g    702 t
ORIGIN

  Query Match             52.8%;  Score 39.6;  DB 9;  Length 3587;
  Best Local Similarity   80.4%;  Pred. No. 0.26;
  Matches   37;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         30 GCCCAGCGCCCAGACAGAGUGUCCCCACACCCUCCUCUGAGACGCC 75
              || |   ||||||||||||:|:||||||||||:||:|:||||||||
Db          2 GCACGAGGCCCAGACAGAGTGTCCCCACACCCTCCTCTGAGACGCC 47


RESULT 15
AF189287
LOCUS       AF189287                796 bp    DNA     linear   ROD 28-OCT-1999
DEFINITION  Mus musculus zinc finger transcription factor GLI (Gli) gene,
            promoter and 5'UTR, partial sequence.
ACCESSION   AF189287
VERSION     AF189287.1  GI:6137222
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 796)
  AUTHORS   Walterhouse,D., Ahmed,M., Slusarski,D., Kalamaras,J., Boucher,D.,
            Holmgren,R. and Iannaccone,P.
  TITLE     gli, a zinc finger transcription factor and oncogene, is expressed
```

during normal mouse development
JOURNAL Dev. Dyn. 196 (2), 91-102 (1993)
MEDLINE 93372381
PUBMED 8364225
REFERENCE 2 (bases 1 to 796)
AUTHORS Liu,C.Z., Yang,J.T., Yoon,J.W., Villavicencio,E., Pfendler,K., Walterhouse,D. and Iannaccone,P.
TITLE Characterization of the promoter region and genomic organization of GLI, a member of the Sonic hedgehog-Patched signaling pathway
JOURNAL Gene 209 (1-2), 1-11 (1998)
MEDLINE 98192509
PUBMED 9524201
REFERENCE 3 (bases 1 to 796)
AUTHORS Liu,C.Z., Yang,J.T., Yoon,J.W., Villavicencio,E., Pfendler,K., Walterhouse,D. and Iannaccone,P.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) Pediatrics, Children's Memorial Institute for Education and Research, 2430 N. Halsted St., CMIER C-503, Chicago, IL 60614, USA

FEATURES	Location/Qualifiers
source	1. .796
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="10"
gene	1. .>796
	/gene="Gli"
promoter	1. .241
	/gene="Gli"
mRNA	242. .>796
	/gene="Gli"
	/product="zinc finger transcription factor GLI"
5'UTR	242. .>796
	/gene="Gli"

BASE COUNT 152 a 215 c 261 g 168 t
ORIGIN

Query Match 49.1%; Score 36.8; DB 10; Length 796;
Best Local Similarity 79.2%; Pred. No. 2.2;
Matches 38; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||  :|||||||:|||||  ||||:|||||||  |  ||||||  |||  |||
Db        330 AGTTTCCAGCCCTGGACCACGCATCCCGAGCACCGCGCCCCGACGGAG 377
```

Search completed: November 13, 2003, 13:18:13
Job time : 1420 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 12:45:23 ; Search time 254 Seconds
(without alignments)
797.078 Million cell updates/sec

Title: US-09-880-253B-5
Perfect score: 75
Sequence: 1 agacuccagcccuggaccgc.....acacccuccucugagacgcc 75

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	75	100.0	75	24	ABK30106	Human GLI gene, Hg
2	75	100.0	3600	22	AAD12302	Human Cubitus inte
3	75	100.0	3600	24	AAL45542	Human Gli1 coding
4	75	100.0	3600	24	AAL45543	Human Gli1 coding
5	75	100.0	3600	24	AAL45544	Human Gli1 coding
6	75	100.0	3600	24	AAL45545	Human Gli1 coding
7	75	100.0	3600	24	ABK30501	Human glioma-assoc
8	73.4	97.9	75	24	ABK30157	Human Hgamma-UTR p
9	71.8	95.7	75	24	ABK30158	Huma Hgamma-UTR pl
10	70.2	93.6	75	24	ABK30159	Human Hbgamma-UTR
11	50.2	66.9	219	24	ABK30105	Human GLI gene, Hb
12	48.6	64.8	219	24	ABK30154	Human Hbeta-UTR pl
13	48	64.0	581	24	ABK30136	Human GLI UTR part
14	48	64.0	1492	24	ABK30508	Human glioma-assoc
15	48	64.0	4620	24	ABK30161	Human GLI1 genomic
16	47	62.7	219	24	ABK30155	Huma Hbeta-UTR plu
17	47	62.7	219	24	ABK30156	Human Hbeta-UTR pl
18	45.4	60.5	74	24	ABK30104	Mouse GLI gene, Mg
19	43.8	58.4	74	24	ABK30151	Mouse Mgamma-UTR p
20	42.2	56.3	74	24	ABK30152	Mouse Mgamma-UTR p
21	40.6	54.1	74	24	ABK30153	Mouse Mgamma-UTR p
22	36.8	49.1	188	24	ABK30103	Mouse GLI gene, Mb
23	36.8	49.1	307	24	ABK30102	Mouse GLI gene, Ma
24	36.8	49.1	307	24	ABK30142	Mouse Malpha-UTR.
25	36.8	49.1	877	24	ABK30134	Mouse GLI UTR part
26	36.8	49.1	3707	24	ABK30160	Mouse GLI1 genomic
27	35.2	46.9	188	24	ABK30148	Mouse Mbeta-UTR pl
c 28	34.2	45.6	8513	22	AAS45355	Chemically pretrea
c 29	34.2	45.6	8513	22	AAS46368	Tumour suppressor
c 30	34.2	45.6	8513	24	ABN80095	Human chemically m
c 31	34.2	45.6	8513	24	ABK28188	DNA transcription
32	33.6	44.8	188	24	ABK30149	Mouse Mbeta-UTR pl
33	33.6	44.8	188	24	ABK30150	Mouse Mbeta-UTR pl
34	33.6	44.8	307	24	ABK30143	Mouse Malpha-UTR p
35	30.4	40.5	307	24	ABK30144	Mouse Malpha-UTR p
c 36	29.8	39.7	2080	24	AAD22007	Human transporters
37	29.6	39.5	770	21	AAC70773	Single nucleotide
38	29.6	39.5	770	21	AAC70782	Single nucleotide
39	29.6	39.5	770	21	AAC70785	Single nucleotide
40	29.6	39.5	770	21	AAC70791	Single nucleotide
c 41	29.2	38.9	1620	24	ABA99034	Human G protein-co
42	29	38.7	573	24	ABK30137	Human GLI UTR part
43	29	38.7	33654	22	AAF70259	Human dopamine rec
44	29	38.7	33654	25	ABT34216	Human dopamine rec
45	28.8	38.4	307	24	ABK30145	Mouse Malpha-UTR p

ALIGNMENTS

RESULT 1
ABK30106
ID ABK30106 standard; RNA; 75 BP.
XX
AC ABK30106;
XX
DT 23-APR-2002 (first entry)
XX
DE Human GLI gene, Hgamma-UTR mRNA.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site;
KW herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves

```
PT   introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT   of authentic site  -
XX
PS   Claim 51; Page 103; 147pp; English.
XX
CC   The invention relates to a method of modulating expression of a genetic
CC   sequence, comprising introducing, creating or deleting one or more
CC   pseudo-translation initiation sites, in the nucleotide sequence of an
CC   mRNA, 5' upstream of the authentic translation initiation site of an open
CC   reading frame (ORF), or by introducing, creating or deleting Kozac
CC   sequences genetically proximal to the pseudo-translation initiation
CC   sites.  The method is useful for modulating the expression of a target
CC   genetic sequence. The method is useful for gene therapy applications and
CC   for expressing traits (herbicide and pesticide resistance) at selective
CC   levels in plants. The genetic constructs are useful for administration
CC   to modulate the expression of an antigen. The method is also useful for
CC   the generation of a genetically modified monocotyledon or dicotyledon
CC   plants, and also for upregulating or downregulating the function of a
CC   promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC   and PCR primers of the invention.
XX
SQ   Sequence 75 BP; 15 A; 35 C; 17 G; 8 U; 0 other;

  Query Match             100.0%;  Score 75;  DB 24;  Length 75;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-13;
  Matches   75;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60

Qy         61 CUCCUCUGAGACGCC 75
              |||||||||||||||
Db         61 CUCCUCUGAGACGCC 75


RESULT 2
AAD12302
ID   AAD12302 standard; cDNA; 3600 BP.
XX
AC   AAD12302;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human Cubitus interruptus (Ci) homologue, GLI-1 cDNA.
XX
KW   Human; transgenic non-human animal; Cubitus interruptus; Ci; GLI-1;
KW   basal cell carcinoma; BCC model system; tumour; screening;
KW   anti-cancer; trichoepithelioma; cylindroma; trichoblastoma; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..3399
FT                   /*tag=  a
FT                   /product= "Human Ci homologue, GLI-1"
XX
PN   WO200156376-A1.
XX
PD   09-AUG-2001.
XX
PF   02-FEB-2001; 2001WO-SE00204.
XX
PR   03-FEB-2000; 2000SE-0000345.
XX
PA   (KARO-) KAROLINSKA INNOVATIONS AB.
XX
PI   Toftgard R;
XX
DR   WPI; 2001-488828/53.
DR   P-PSDB; AAE06644.
XX
PT   Transgenic non-human animal useful as basal cell carcinoma model system
PT   to identify anti-cancer drug candidates, overexpresses transgene
PT   encoding GLI-1 protein which is a human homolog to Cubitus interruptus
PT   -
XX
PS   Claim 6; Page 25-26; 33pp; English.
XX
CC   The present invention relates to a transgenic non-human animal
CC   comprising a transgene containing a nucleic acid encoding a human
CC   Cubitus interruptus (Ci) homologue protein, GLI-1. The transgenic
CC   non-human animal is useful as basal cell carcinoma (BCC) model system
CC   since it overexpresses GLI-1 which leads to development of tumours
CC   resembling human BCC. Thus it is also useful for screening anti-cancer
CC   drug candidates and evaluating whether it affects BCC,
CC   trichoepitheliomas, cylindromas and trichoblastomas. The present
CC   sequence is a cDNA encoding human Ci homologue protein, GLI-1.
XX
SQ   Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 other;

  Query Match             100.0%;  Score 75;  DB 22;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 1.8e-13;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||:|:||||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 3
AAL45542
ID   AAL45542 standard; cDNA; 3600 BP.
XX
AC   AAL45542;
XX
DT   11-JUN-2002  (first entry)
XX
DE   Human Gli1 coding sequence SEQ ID NO: 12.
XX
KW   Gli1; screening method; bone induction; cartilage induction;
KW   orthopaedic disease; dental disease; osteoporosis; hyperosteogenesis;
KW   osteopathic; antiarthritic; vulnerary; immunosuppressive; human;
KW   hyperchondrogenesis; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..3399
FT                   /*tag=  a
FT                   /product= "Gli1"
XX
PN   WO200211752-A1.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-JP06688.
XX
PR   04-AUG-2000; 2000JP-0242767.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Hikichi Y;
XX
DR   WPI; 2002-241709/29.
DR   P-PSDB; AAO17109.
XX
PT   Promotion of bone and cartilage formation using Gli1 protein or DNA
PT   encoding it for treatment of skeletal disorders  -
XX
PS   Claim 6; Page 100-101; 154pp; Japanese.
XX
```

CC The present invention relates to agents for the promotion of bone and
CC cartilage formation which contain as the active component a Gli1 protein
CC or a DNA encoding a Gli1 protein. The agents can be used in the
CC prevention, treatment and diagnosis of bone and cartilage disorders
CC including bone fractures, joint deformation, osteoarthritis,
CC osteoporosis, cartilage damage, trauma, bone formation defects, cartilage
CC formation defects, bone defects, dental disease, hyperosteogenesis and
CC hyperchondrogenesis, and for use in cosmetic and therapeutic bone
CC transplantation. The present sequence is a human Gli1 coding sequence
CC described in the exemplification of the invention.
XX
SQ Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 other;

Query Match 100.0%; Score 75; DB 24; Length 3600;
Best Local Similarity 89.3%; Pred. No. 1.8e-13;
Matches 67; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:|||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78
```

RESULT 4
AAL45543
ID AAL45543 standard; cDNA; 3600 BP.
XX
AC AAL45543;
XX
DT 11-JUN-2002 (first entry)
XX
DE Human Gli1 coding sequence SEQ ID NO: 14.
XX
KW Gli1; screening method; bone induction; cartilage induction;
KW orthopaedic disease; dental disease; osteoporosis; hyperosteogenesis;
KW osteopathic; antiarthritic; vulnerary; immunosuppressive; human;
KW hyperchondrogenesis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 79..3399
FT /*tag= a
FT /product= "Gli1"
XX
PN WO200211752-A1.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-JP06688.
XX
PR 04-AUG-2000; 2000JP-0242767.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Hikichi Y;
XX
DR WPI; 2002-241709/29.
DR P-PSDB; AAO17110.
XX
PT Promotion of bone and cartilage formation using Gli1 protein or DNA
PT encoding it for treatment of skeletal disorders -
XX
PS Claim 6; Page 105-109; 154pp; Japanese.
XX
CC The present invention relates to agents for the promotion of bone and
CC cartilage formation which contain as the active component a Gli1 protein
CC or a DNA encoding a Gli1 protein. The agents can be used in the
CC prevention, treatment and diagnosis of bone and cartilage disorders
CC including bone fractures, joint deformation, osteoarthritis,
CC osteoporosis, cartilage damage, trauma, bone formation defects, cartilage
CC formation defects, bone defects, dental disease, hyperosteogenesis and
CC hyperchondrogenesis, and for use in cosmetic and therapeutic bone
CC transplantation. The present sequence is a human Gli1 coding sequence
CC described in the exemplification of the invention.
XX
SQ Sequence 3600 BP; 784 A; 1162 C; 949 G; 705 T; 0 other;

Query Match 100.0%; Score 75; DB 24; Length 3600;
Best Local Similarity 89.3%; Pred. No. 1.8e-13;
Matches 67; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:|||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78
```

RESULT 5
AAL45544
ID AAL45544 standard; cDNA; 3600 BP.
XX
AC AAL45544;
XX
DT 11-JUN-2002 (first entry)
XX
DE Human Gli1 coding sequence SEQ ID NO: 16.
XX
KW Gli1; screening method; bone induction; cartilage induction;
KW orthopaedic disease; dental disease; osteoporosis; hyperosteogenesis;
KW osteopathic; antiarthritic; vulnerary; immunosuppressive; human;
KW hyperchondrogenesis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 79..3399
FT /*tag= a
FT /product= "Gli1"
FT /transl_except= (pos:2875..2877,aa:Glu)
XX
PN WO200211752-A1.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-JP06688.
XX
PR 04-AUG-2000; 2000JP-0242767.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Hikichi Y;
XX
DR WPI; 2002-241709/29.
DR P-PSDB; AAO17111.
XX
PT Promotion of bone and cartilage formation using Gli1 protein or DNA
PT encoding it for treatment of skeletal disorders -
XX
PS Claim 6; Page 115-117; 154pp; Japanese.
XX
CC The present invention relates to agents for the promotion of bone and
CC cartilage formation which contain as the active component a Gli1 protein
CC or a DNA encoding a Gli1 protein. The agents can be used in the
CC prevention, treatment and diagnosis of bone and cartilage disorders
CC including bone fractures, joint deformation, osteoarthritis,
CC osteoporosis, cartilage damage, trauma, bone formation defects, cartilage
CC formation defects, bone defects, dental disease, hyperosteogenesis and
CC hyperchondrogenesis, and for use in cosmetic and therapeutic bone
CC transplantation. The present sequence is a human Gli1 coding sequence

```
CC   described in the exemplification of the invention.
XX
SQ   Sequence 3600 BP; 786 A; 1161 C; 948 G; 705 T; 0 other;

  Query Match             100.0%;  Score 75;  DB 24;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 1.8e-13;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 6
AAL45545
ID   AAL45545 standard; cDNA; 3600 BP.
XX
AC   AAL45545;
XX
DT   11-JUN-2002  (first entry)
XX
DE   Human Gli1 coding sequence SEQ ID NO: 18.
XX
KW   Gli1; screening method; bone induction; cartilage induction;
KW   orthopaedic disease; dental disease; osteoporosis; hyperosteogenesis;
KW   osteopathic; antiarthritic; vulnerary; immunosuppressive; human;
KW   hyperchondrogenesis; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..3399
FT                   /*tag=  a
FT                   /product= "Gli1"
FT                   /transl_except= (pos:2416..2599,aa:Trp)
FT                   /note= "there is an apparent 180 nucleotide
FT                   insertion between residues 780 and 781 of the Gli1
FT                   protein"
XX
PN   WO200211752-A1.
XX
PD   14-FEB-2002.
XX
PF   03-AUG-2001; 2001WO-JP06688.
XX
PR   04-AUG-2000; 2000JP-0242767.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Hikichi Y;
XX
DR   WPI; 2002-241709/29.
DR   P-PSDB; AAO17112.
XX
PT   Promotion of bone and cartilage formation using Gli1 protein or DNA
PT   encoding it for treatment of skeletal disorders  -
XX
PS   Claim 6; Page 123-125; 154pp; Japanese.
XX
CC   The present invention relates to agents for the promotion of bone and
CC   cartilage formation which contain as the active component a Gli1 protein
CC   or a DNA encoding a Gli1 protein. The agents can be used in the
CC   prevention, treatment and diagnosis of bone and cartilage disorders
CC   including bone fractures, joint deformation, osteoarthritis,
CC   osteoporosis, cartilage damage, trauma, bone formation defects, cartilage
CC   formation defects, bone defects, dental disease, hyperosteogenesis and
CC   hyperchondrogenesis, and for use in cosmetic and therapeutic bone
CC   transplantation. The present sequence is a human Gli1 coding sequence
CC   described in the exemplification of the invention.
XX
SQ   Sequence 3600 BP; 785 A; 1162 C; 948 G; 705 T; 0 other;

  Query Match             100.0%;  Score 75;  DB 24;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 1.8e-13;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:||||||||||||||||||||||||:|:|||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:||||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 7
ABK30501
ID   ABK30501 standard; DNA; 3600 BP.
XX
AC   ABK30501;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human glioma-associated oncogene-1 DNA sequence.
XX
KW   Human; glioma-associated oncogene-1 associated disease; infection;
KW   inflammation; tumour formation; cytostatic; antiinflammatory;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..3399
FT                   /*tag=  a
FT                   /product= "Glioma-associated oncogene-1 protein"
XX
PN   US6329203-B1.
XX
PD   11-DEC-2001.
XX
PF   08-SEP-2000; 2000US-0657042.
XX
PR   08-SEP-2000; 2000US-0657042.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Wyatt J;
XX
DR   WPI; 2002-138363/18.
DR   P-PSDB; AAU12085.
XX
PT   Novel antisense compounds targeted to nucleic acids encoding
PT   glioma-associated oncogene-1, for modulating the gene expression and
PT   treating diseases associated with expression of the oncogene in humans
PT   -
XX
PS   Example 13; Column 47-56; 43pp; English.
XX
CC   The present invention relates to antisense compounds and methods for
CC   modulating the expression of human glioma-associated oncogene-1. The
CC   antisense compounds, particularly antisense oligonucleotides
CC   (ABK30509-ABK30586), target and inhibit the expression of human
CC   glioma-associated oncogene-1. The antisense compounds are useful for
CC   inhibiting the expression of human glioma-associated oncogene-1 in human
CC   cells or tissues and for treating an animal, particularly a human
CC   suspected of having or being prone to a disease or condition associated
CC   with expression of glioma-associated oncogene-1. The compounds are useful
CC   for diagnostics, therapeutics and as research reagent,
CC   e.g. prophylactically to prevent or delay infection, inflammation or
CC   tumour formation. The antisense compounds are safely and effectively
CC   administered to humans. The present sequence represents human
CC   glioma-associated oncogene-1 DNA.
```

```
XX
SQ   Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 other;

  Query Match             100.0%;  Score 75;  DB 24;  Length 3600;
  Best Local Similarity   89.3%;  Pred. No. 1.8e-13;
  Matches   67;  Conservative    8;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              ||||:|||||||:||||||||||:|||||||||||||||||||||||||:|:||||||||||
Db          4 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAGTGTCCCCACACC 63

Qy         61 CUCCUCUGAGACGCC 75
              |:||:|:|||||||
Db         64 CTCCTCTGAGACGCC 78


RESULT 8
ABK30157
ID   ABK30157 standard; RNA; 75 BP.
XX
AC   ABK30157;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human Hgamma-UTR plus 1 ATG translation initiation site.
XX
KW   Human; mouse; gene therapy; pseudo-translation initiation site;
KW   herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS   Homo sapiens.
XX
PN   WO200196569-A1.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-AU00697.
XX
PR   13-JUN-2000; 2000US-211159P.
XX
PA   (UYQU ) UNIV QUEENSLAND.
XX
PI   Rothnagel JA,  Wang X;
XX
DR   WPI; 2002-098072/13.
XX
PT   Modulating expression of genetic sequence, comprising ORF having
PT   RTG/RUG corresponding to authentic translation site, involves
PT   introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT   of authentic site  -
XX
PS   Example 30; Page 75; 147pp; English.
XX
CC   The invention relates to a method of modulating expression of a genetic
CC   sequence, comprising introducing, creating or deleting one or more
CC   pseudo-translation initiation sites, in the nucleotide sequence of an
CC   mRNA, 5' upstream of the authentic translation initiation site of an open
CC   reading frame (ORF), or by introducing, creating or deleting Kozac
CC   sequences genetically proximal to the pseudo-translation initiation
CC   sites.  The method is useful for modulating the expression of a target
CC   genetic sequence. The method is useful for gene therapy applications and
CC   for expressing traits (herbicide and pesticide resistance) at selective
CC   levels in plants. The genetic constructs are useful for administration
CC   to modulate the expression of an antigen. The method is also useful for
CC   the generation of a genetically modified monocotyledon or dicotyledon
CC   plants, and also for upregulating or downregulating the function of a
CC   promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC   and PCR primers of the invention.
XX
SQ   Sequence 75 BP; 16 A; 34 C; 17 G; 8 U; 0 other;

  Query Match             97.9%;  Score 73.4;  DB 24;  Length 75;
  Best Local Similarity   98.7%;  Pred. No. 4.3e-13;
  Matches   74;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACUCCAGCCAUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60

Qy         61 CUCCUCUGAGACGCC 75
              ||||||||||||||||
Db         61 CUCCUCUGAGACGCC 75


RESULT 9
ABK30158
ID   ABK30158 standard; RNA; 75 BP.
XX
AC   ABK30158;
XX
DT   23-APR-2002  (first entry)
XX
DE   Huma Hgamma-UTR plus 2 ATG translation initiation site.
XX
KW   Human; mouse; gene therapy; pseudo-translation initiation site;
KW   herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS   Homo sapiens.
XX
PN   WO200196569-A1.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-AU00697.
XX
PR   13-JUN-2000; 2000US-211159P.
XX
PA   (UYQU ) UNIV QUEENSLAND.
XX
PI   Rothnagel JA,  Wang X;
XX
DR   WPI; 2002-098072/13.
XX
PT   Modulating expression of genetic sequence, comprising ORF having
PT   RTG/RUG corresponding to authentic translation site, involves
PT   introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT   of authentic site  -
XX
PS   Example 31; Page 75; 147pp; English.
XX
CC   The invention relates to a method of modulating expression of a genetic
CC   sequence, comprising introducing, creating or deleting one or more
CC   pseudo-translation initiation sites, in the nucleotide sequence of an
CC   mRNA, 5' upstream of the authentic translation initiation site of an open
CC   reading frame (ORF), or by introducing, creating or deleting Kozac
CC   sequences genetically proximal to the pseudo-translation initiation
CC   sites.  The method is useful for modulating the expression of a target
CC   genetic sequence. The method is useful for gene therapy applications and
CC   for expressing traits (herbicide and pesticide resistance) at selective
CC   levels in plants. The genetic constructs are useful for administration
CC   to modulate the expression of an antigen. The method is also useful for
CC   the generation of a genetically modified monocotyledon or dicotyledon
CC   plants, and also for upregulating or downregulating the function of a
CC   promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC   and PCR primers of the invention.
XX
SQ   Sequence 75 BP; 16 A; 33 C; 18 G; 8 U; 0 other;

  Query Match             95.7%;  Score 71.8;  DB 24;  Length 75;
  Best Local Similarity   97.3%;  Pred. No. 1.3e-12;
  Matches   73;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||||||| |||||||||||  |||||||||||||||||||||||||||||||||||||
Db          1 AGACUCCAGCCAUGGACCGCGCAUGCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60

Qy         61 CUCCUCUGAGACGCC 75
```

```
                  ||||||||||||||||
Db          61 CUCCUCUGAGACGCC 75
```

RESULT 10
ABK30159
ID ABK30159 standard; RNA; 75 BP.
XX
AC ABK30159;
XX
DT 23-APR-2002 (first entry)
XX
DE Human Hbgamma-UTR plus 3 ATG translation initiation site.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site;
KW herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves
PT introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT of authentic site -
XX
PS Example 32; Page 76; 147pp; English.
XX
CC The invention relates to a method of modulating expression of a genetic
CC sequence, comprising introducing, creating or deleting one or more
CC pseudo-translation initiation sites, in the nucleotide sequence of an
CC mRNA, 5' upstream of the authentic translation initiation site of an open
CC reading frame (ORF), or by introducing, creating or deleting Kozac
CC sequences genetically proximal to the pseudo-translation initiation
CC sites. The method is useful for modulating the expression of a target
CC genetic sequence. The method is useful for gene therapy applications and
CC for expressing traits (herbicide and pesticide resistance) at selective
CC levels in plants. The genetic constructs are useful for administration
CC to modulate the expression of an antigen. The method is also useful for
CC the generation of a genetically modified monocotyledon or dicotyledon
CC plants, and also for upregulating or downregulating the function of a
CC promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC and PCR primers of the invention.
XX
SQ Sequence 75 BP; 16 A; 32 C; 18 G; 9 U; 0 other;

```
  Query Match             93.6%;  Score 70.2;  DB 24;  Length 75;
  Best Local Similarity   96.0%;  Pred. No. 3.9e-12;
  Matches   72;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||
Db          1 AGACUCCAGCCAUGGACCGCGCAUGCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60

Qy         61 CUCCUCUGAGACGCC 75
              ||||||||||||| |||
Db         61 CUCCUCUGAGAUGCC 75
```

RESULT 11
ABK30105
ID ABK30105 standard; RNA; 219 BP.
XX
AC ABK30105;
XX
DT 23-APR-2002 (first entry)
XX
DE Human GLI gene, Hbeta-UTR mRNA.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site;
KW herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves
PT introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT of authentic site -
XX
PS Claim 50; Page 103; 147pp; English.
XX
CC The invention relates to a method of modulating expression of a genetic
CC sequence, comprising introducing, creating or deleting one or more
CC pseudo-translation initiation sites, in the nucleotide sequence of an
CC mRNA, 5' upstream of the authentic translation initiation site of an open
CC reading frame (ORF), or by introducing, creating or deleting Kozac
CC sequences genetically proximal to the pseudo-translation initiation
CC sites. The method is useful for modulating the expression of a target
CC genetic sequence. The method is useful for gene therapy applications and
CC for expressing traits (herbicide and pesticide resistance) at selective
CC levels in plants. The genetic constructs are useful for administration
CC to modulate the expression of an antigen. The method is also useful for
CC the generation of a genetically modified monocotyledon or dicotyledon
CC plants, and also for upregulating or downregulating the function of a
CC promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC and PCR primers of the invention.
XX
SQ Sequence 219 BP; 44 A; 65 C; 57 G; 53 U; 0 other;

```
  Query Match             66.9%;  Score 50.2;  DB 24;  Length 219;
  Best Local Similarity   81.7%;  Pred. No. 4.3e-06;
  Matches   58;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| |    |   |
Db          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUCUGUGUAUCUC 60

Qy         61 CUCCUCUGAGA 71
                 ||| | ||
Db         61 UGUCUCAGGGA 71
```

RESULT 12
ABK30154
ID ABK30154 standard; RNA; 219 BP.
XX
AC ABK30154;
XX
DT 23-APR-2002 (first entry)
XX

DE Human Hbeta-UTR plus 1 ATG translation initiation site.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site;
KW herbicide resistance; pesticide resistance; transgenic plant; ss.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves
PT introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT of authentic site -
XX
PS Example 27; Page 74; 147pp; English.
XX
CC The invention relates to a method of modulating expression of a genetic
CC sequence, comprising introducing, creating or deleting one or more
CC pseudo-translation initiation sites, in the nucleotide sequence of an
CC mRNA, 5' upstream of the authentic translation initiation site of an open
CC reading frame (ORF), or by introducing, creating or deleting Kozac
CC sequences genetically proximal to the pseudo-translation initiation
CC sites. The method is useful for modulating the expression of a target
CC genetic sequence. The method is useful for gene therapy applications and
CC for expressing traits (herbicide and pesticide resistance) at selective
CC levels in plants. The genetic constructs are useful for administration
CC to modulate the expression of an antigen. The method is also useful for
CC the generation of a genetically modified monocotyledon or dicotyledon
CC plants, and also for upregulating or downregulating the function of a
CC promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC and PCR primers of the invention.
XX
SQ Sequence 219 BP; 45 A; 64 C; 57 G; 53 U; 0 other;

Query Match 64.8%; Score 48.6; DB 24; Length 219;
Best Local Similarity 80.3%; Pred. No. 1.3e-05;
Matches 57; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||||||||||  |||||||||||||||||||||||||||||||||||| |    |   |
Db          1 AGACUCCAGCCAUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUCUGUGUAUCUC 60

Qy         61 CUCCUCUGAGA 71
                 |||  | ||
Db         61 UGUCUCAGGGA 71
```

RESULT 13
ABK30136
ID ABK30136 standard; DNA; 581 BP.
XX
AC ABK30136;
XX
DT 23-APR-2002 (first entry)
XX
DE Human GLI UTR partial genomic sequence #1.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site; gene;
KW herbicide resistance; pesticide resistance; transgenic plant; ds.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves
PT introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT of authentic site -
XX
PS Example 7; Page 111; 147pp; English.
XX
CC The invention relates to a method of modulating expression of a genetic
CC sequence, comprising introducing, creating or deleting one or more
CC pseudo-translation initiation sites, in the nucleotide sequence of an
CC mRNA, 5' upstream of the authentic translation initiation site of an open
CC reading frame (ORF), or by introducing, creating or deleting Kozac
CC sequences genetically proximal to the pseudo-translation initiation
CC sites. The method is useful for modulating the expression of a target
CC genetic sequence. The method is useful for gene therapy applications and
CC for expressing traits (herbicide and pesticide resistance) at selective
CC levels in plants. The genetic constructs are useful for administration
CC to modulate the expression of an antigen. The method is also useful for
CC the generation of a genetically modified monocotyledon or dicotyledon
CC plants, and also for upregulating or downregulating the function of a
CC promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC and PCR primers of the invention.
XX
SQ Sequence 581 BP; 114 A; 126 C; 196 G; 145 T; 0 other;

Query Match 64.0%; Score 48; DB 24; Length 581;
Best Local Similarity 93.8%; Pred. No. 2.1e-05;
Matches 45; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
              ||||:|||||||:||||||||||:||||||||||||||||||||||||
Db          1 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 48
```

RESULT 14
ABK30508
ID ABK30508 standard; DNA; 1492 BP.
XX
AC ABK30508;
XX
DT 23-APR-2002 (first entry)
XX
DE Human glioma-associated oncogene-1 partial DNA sequence.
XX
KW Human; glioma-associated oncogene-1 associated disease; infection;
KW inflammation; tumour formation; cytostatic; antiinflammatory; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1017..1067
FT /*tag= a
FT /partial
FT /product= "Partial peptide for human glioma-associated
FT oncogene-1"
FT /note= "This sequence lacks both start and stop codons"
XX
PN US6329203-B1.

XX
PD 11-DEC-2001.
XX
PF 08-SEP-2000; 2000US-0657042.
XX
PR 08-SEP-2000; 2000US-0657042.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Wyatt J;
XX
DR WPI; 2002-138363/18.
DR P-PSDB; AAU12086.
XX
PT Novel antisense compounds targeted to nucleic acids encoding
PT glioma-associated oncogene-1, for modulating the gene expression and
PT treating diseases associated with expression of the oncogene in humans
PT -
XX
PS Example 15; Column 57-60; 43pp; English.
XX
CC The present invention relates to antisense compounds and methods for
CC modulating the expression of human glioma-associated oncogene-1. The
CC antisense compounds, particularly antisense oligonucleotides
CC (ABK30509-ABK30586), target and inhibit the expression of human
CC glioma-associated oncogene-1. The antisense compounds are useful for
CC inhibiting the expression of human glioma-associated oncogene-1 in human
CC cells or tissues and for treating an animal, particularly a human
CC suspected of having or being prone to a disease or condition associated
CC with expression of glioma-associated oncogene-1. The compounds are useful
CC for diagnostics, therapeutics and as research reagent,
CC e.g. prophylactically to prevent or delay infection, inflammation or
CC tumour formation. The antisense compounds are safely and effectively
CC administered to humans. The present sequence represents a partial
CC DNA sequence for human glioma-associated oncogene-1.
XX
SQ Sequence 1492 BP; 293 A; 415 C; 471 G; 313 T; 0 other;

Query Match 64.0%; Score 48; DB 24; Length 1492;
Best Local Similarity 93.8%; Pred. No. 2.2e-05;
Matches 45; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
||||:|||||||:||||||||||:||||||||||||||||||||||
Db 1020 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 1067

RESULT 15
ABK30161
ID ABK30161 standard; DNA; 4620 BP.
XX
AC ABK30161;
XX
DT 23-APR-2002 (first entry)
XX
DE Human GLI1 genomic sequence.
XX
KW Human; mouse; gene therapy; pseudo-translation initiation site; gene;
KW herbicide resistance; pesticide resistance; transgenic plant; ds.
XX
OS Homo sapiens.
XX
PN WO200196569-A1.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-AU00697.
XX
PR 13-JUN-2000; 2000US-211159P.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Rothnagel JA, Wang X;
XX
DR WPI; 2002-098072/13.
XX
PT Modulating expression of genetic sequence, comprising ORF having
PT RTG/RUG corresponding to authentic translation site, involves
PT introducing/removing RTG/RUG triplets in nucleotide sequence upstream
PT of authentic site -
XX
PS Example 7; Page 120-122; 147pp; English.
XX
CC The invention relates to a method of modulating expression of a genetic
CC sequence, comprising introducing, creating or deleting one or more
CC pseudo-translation initiation sites, in the nucleotide sequence of an
CC mRNA, 5' upstream of the authentic translation initiation site of an open
CC reading frame (ORF), or by introducing, creating or deleting Kozac
CC sequences genetically proximal to the pseudo-translation initiation
CC sites. The method is useful for modulating the expression of a target
CC genetic sequence. The method is useful for gene therapy applications and
CC for expressing traits (herbicide and pesticide resistance) at selective
CC levels in plants. The genetic constructs are useful for administration
CC to modulate the expression of an antigen. The method is also useful for
CC the generation of a genetically modified monocotyledon or dicotyledon
CC plants, and also for upregulating or downregulating the function of a
CC promoter. ABK30102-ABK30161 represent human and mouse GLI gene sequences
CC and PCR primers of the invention.
XX
SQ Sequence 4620 BP; 916 A; 1354 C; 1368 G; 982 T; 0 other;

Query Match 64.0%; Score 48; DB 24; Length 4620;
Best Local Similarity 93.8%; Pred. No. 2.4e-05;
Matches 45; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAG 48
||||:|||||||:||||||||||:||||||||||||||||||||||
Db 1034 AGACTCCAGCCCTGGACCGCGCATCCCGAGCCCAGCGCCCAGACAGAG 1081

Search completed: November 13, 2003, 12:54:20
Job time : 254 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 12:47:43 ; Search time 2051 Seconds
(without alignments)
888.755 Million cell updates/sec

Title: US-09-880-253B-5
Perfect score: 75
Sequence: 1 agacuccagcccuggaccgc.....acacccuccucugagacgcc 75

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	39	52.0	767	12	BI114524	BI114524 602862031
	2	31	41.3	1008	13	BQ669921	BQ669921 AGENCOURT
c	3	30.6	40.8	364	14	R91407	R91407 yq04h02.s1
c	4	30.4	40.5	812	13	BU424864	BU424864 603230655
c	5	29.8	39.7	574	12	BI824472	BI824472 603038839
c	6	29.8	39.7	575	12	BI547637	BI547637 603191773
c	7	29.8	39.7	641	12	BI517462	BI517462 603041736
c	8	29.8	39.7	669	12	BI757775	BI757775 603029745
c	9	29.8	39.7	715	12	BI755911	BI755911 603030127
c	10	29.8	39.7	1080	13	BX342864	BX342864 BX342864
c	11	29.8	39.7	1201	13	BX397064	BX397064 BX397064
c	12	29.6	39.5	407	12	BI724568	BI724568 1031072H0
c	13	29.6	39.5	516	12	BM003088	BM003088 1031108E0
	14	29.4	39.2	563	29	BZ639961	BZ639961 OGCCP26TC
c	15	29.4	39.2	867	10	BF616342	BF616342 HVSMEc000
c	16	29.4	39.2	1000	13	BX342523	BX342523 BX342523
	17	29.2	38.9	456	13	BX118143	BX118143 BX118143
	18	29	38.7	390	9	AI123001	AI123001 ok36c01.s
c	19	29	38.7	393	10	BG027933	BG027933 602294531
	20	29	38.7	745	12	BG827135	BG827135 602751117
	21	29	38.7	780	12	BI195997	BI195997 602756310
	22	28.8	38.4	265	10	BB585305	BB585305 BB585305
	23	28.8	38.4	461	12	BM795944	BM795944 K-EST0078
	24	28.8	38.4	463	10	BE394194	BE394194 601311694
c	25	28.8	38.4	500	13	BU584167	BU584167 2074149T7
	26	28.8	38.4	511	9	AI692466	AI692466 wd70g06.x
c	27	28.8	38.4	530	9	AW798307	AW798307 RC1-UM004
	28	28.8	38.4	576	12	BM700772	BM700772 UI-E-CL1-
	29	28.8	38.4	582	13	BU584165	BU584165 2074149F7
	30	28.8	38.4	666	10	BE394361	BE394361 601311309
	31	28.8	38.4	708	13	BX112900	BX112900 BX112900
	32	28.8	38.4	730	12	BM044240	BM044240 603621502
	33	28.8	38.4	767	10	BG750003	BG750003 602708758
	34	28.8	38.4	877	13	BQ720868	BQ720868 AGENCOURT
	35	28.8	38.4	916	12	BM019566	BM019566 603647786
	36	28.8	38.4	938	12	BI093058	BI093058 602858150
	37	28.8	38.4	943	13	BQ945928	BQ945928 AGENCOURT
	38	28.8	38.4	1034	11	BC013276	BC013276 Homo sapi
c	39	28.6	38.1	320	9	AA024106	AA024106 mh98e05.r
	40	28.4	37.9	848	29	AG138174	AG138174 Pan trogl
	41	28.4	37.9	1040	12	BM477912	BM477912 AGENCOURT
c	42	28.2	37.6	389	12	BI261436	BI261436 602953470
c	43	28.2	37.6	561	14	CA334682	CA334682 NISC_1s12
c	44	28.2	37.6	601	12	BM041953	BM041953 603615827
	45	28.2	37.6	968	13	BQ681918	BQ681918 AGENCOURT

ALIGNMENTS

RESULT 1
BI114524
LOCUS BI114524 767 bp mRNA linear EST 26-JUN-2001
DEFINITION 602862031F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5021359 5', mRNA sequence.
ACCESSION BI114524
VERSION BI114524.1 GI:14565425
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 767)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1839 row: m column: 08
High quality sequence stop: 594.

```
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5021359"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      186 a    245 c    201 g    135 t
ORIGIN

  Query Match             52.0%;  Score 39;  DB 12;  Length 767;
  Best Local Similarity   87.2%;  Pred. No. 1;
  Matches   34;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 GCCCAGACAGAGUGUCCCCACACCCUCCUCUGAGACGCC 75
              |||||||||||:|:||||||||||:||:|:||||||||
Db          1 GCCCAGACAGAGTGTCCCCACACCCTCCTCTGAGACGCC 39


RESULT 2
BQ669921
LOCUS       BQ669921               1008 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8208376 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6254737
            5', mRNA sequence.
ACCESSION   BQ669921
VERSION     BQ669921.1  GI:21780755
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1008)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2404  row: 1  column: 02
            High quality sequence stop: 259.
FEATURES             Location/Qualifiers
     source          1. .1008
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6254737"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of  California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      118 a    518 c    219 g    153 t
ORIGIN

  Query Match             41.3%;  Score 31;  DB 13;  Length 1008;
  Best Local Similarity   60.3%;  Pred. No. 1.4e+02;
  Matches   38;  Conservative    5;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          4 CUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUC 63
              | ||  |||:   || | | :|||| |||| ||||||   | |  :| |||| | |||:|
Db        471 CCCCCCCCCTTTCCCCCCCTTCCCGCGCCCCGCGCCCCCCCCGTCTGGCCCCCCGCCCTC 530

Qy         64 CUC 66
              |:|
Db        531 CTC 533


RESULT 3
R91407/c
LOCUS       R91407                 364 bp    mRNA    linear   EST 25-AUG-1995
DEFINITION  yq04h02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:195987 3', mRNA sequence.
ACCESSION   R91407
VERSION     R91407.1  GI:958947
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2298
            High quality sequence stops: 321
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2298   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 321.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3765037"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:195987"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       82 a     79 c     90 g    112 t      1 others
ORIGIN

  Query Match             40.8%;  Score 30.6;  DB 14;  Length 364;
```

```
  Best Local Similarity   60.7%;  Pred. No. 1.5e+02;
  Matches   37;  Conservative    5;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          5 UCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUCC 64
              :|||||||:| || | | | |||   |||| | ||||||   ||  :|||||   ||: |
Db        279 TCCAGCCCTGTACGGAGGACCCCTGACCCATCCCCCAGAAGTAGACTCCCCATCTCCTTC 220

Qy         65 U 65
              :
Db        219 T 219


RESULT 4
BU424864/c
LOCUS       BU424864                812 bp    mRNA    linear   EST 29-NOV-2002
DEFINITION  603230655F1 CSEQRBN09 Gallus gallus cDNA clone ChEST225n14 5', mRNA
            sequence.
ACCESSION   BU424864
VERSION     BU424864.1  GI:25917540
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 812)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .812
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Layer and broiler"
                     /db_xref="taxon:9031"
                     /clone="ChEST225n14"
                     /sex="Male and female"
                     /tissue_type="Chondrocytes isolated from growth plate
                     cartilage"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQRBN09"
                     /note="Vector: pBluescript II KS(+); Site_1: EcoRI;
                     Site_2: NotI; This normalized library was constructed from
                     1 million independent clones. cDNA synthesis was initiated
                     using an oligo(dT) primer, using methylated C in the first
                     strand synthesis reaction. Following this first strand
                     reaction, double-stranded cDNA was blunted, ligated to
                     NotI adapters, digested with EcoRI, size-selected, and
                     cloned into the NotI and EcoRI compatible sites of a
                     custom modified MCS of the pBluescript (KS+) vector. The
                     library was normalized in 2 rounds using conditions
                     adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
                     Bonaldo et al., Genome Research 6 (1996): 791, except that
                     a significantly longer reannealing hybridization was
                     used."
BASE COUNT      164 a    199 c    276 g    173 t
ORIGIN

  Query Match             40.5%;  Score 30.4;  DB 13;  Length 812;
  Best Local Similarity   65.6%;  Pred. No. 2e+02;
  Matches   42;  Conservative    1;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 AGACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACC 60
              |||  | ||||  | || | ||| ||  |||||||| |||| |||||      | ||  ||
Db         80 AGAGCCGAGCCTCGCACTGTGCACCCTCAGCCCAGCACCCATACAGAAGACTCTCATCCC 21

Qy         61 CUCC 64
              |:||
Db         20 CTCC 17


RESULT 5
BI824472/c
LOCUS       BI824472                574 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603038839F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179953 5',
            mRNA sequence.
ACCESSION   BI824472
VERSION     BI824472.1  GI:15936022
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11448  row: m  column: 10
            High quality sequence stop: 568.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5179953"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_115"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
BASE COUNT       68 a    218 c    227 g     61 t
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 12;  Length 574;
  Best Local Similarity   58.9%;  Pred. No. 2.7e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        489 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 430

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        429 TCCTCCCCGGCGC 417


RESULT 6
BI547637/c
```

```
LOCUS       BI547637                575 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603191773F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263002 5',
            mRNA sequence.
ACCESSION   BI547637
VERSION     BI547637.1  GI:15434949
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11663  row: a  column: 19
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5263002"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT       68 a    222 c    230 g     55 t
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 12;  Length 575;
  Best Local Similarity   58.9%;  Pred. No. 2.7e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        490 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 431

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        430 TCCTCCCCGGCGC 418


RESULT 7
BI517462/c
LOCUS       BI517462                641 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603041736F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5162937 5',
            mRNA sequence.
ACCESSION   BI517462
VERSION     BI517462.1  GI:15342254
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11404  row: h  column: 10
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5162937"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_116"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT       76 a    246 c    256 g     63 t
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 12;  Length 641;
  Best Local Similarity   58.9%;  Pred. No. 2.7e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || |.| :| |   ||||| |||| ||||||| | | || |  ||
Db        500 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 441

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        440 TCCTCCCCGGCGC 428


RESULT 8
BI757775/c
LOCUS       BI757775                669 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603029745F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200210 5',
            mRNA sequence.
ACCESSION   BI757775
VERSION     BI757775.1  GI:15749353
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
            http://image.llnl.gov
            Plate: LLAM11501  row: i  column: 11
            High quality sequence stop: 628.
FEATURES              Location/Qualifiers
     source           1. .669
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5200210"
                      /lab_host="DH10B"
                      /clone_lib="NIH_MGC_114"
                      /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                      male brains, age range 23-27 yo. Library is oligo-dT
                      primed and directionally cloned (EcoRV site is  destroyed
                      upon cloning).  Average insert size 1.5 kb, insert size
                      range 1-3 kb.  Library is normalized and enriched for
                      full-length  clones and was constructed by C. Gruber
                      (Invitrogen).  Research  Genetics tracking code 019. Note:
                      this is a NIH_MGC Library."
BASE COUNT       86 a    258 c    256 g     69 t
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 12;  Length 669;
  Best Local Similarity   58.9%;  Pred. No. 2.8e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        503 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 444

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        443 TCCTCCCCGGCGC 431


RESULT 9
BI755911/c
LOCUS       BI755911                715 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603030127F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200494 5',
            mRNA sequence.
ACCESSION   BI755911
VERSION     BI755911.1  GI:15747489
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11502  row: e  column: 07
            High quality sequence stop: 628.
FEATURES              Location/Qualifiers
     source           1. .715
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5200494"
                      /lab_host="DH10B"
                      /clone_lib="NIH_MGC_114"
                      /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                      male brains, age range 23-27 yo. Library is oligo-dT
                      primed and directionally cloned (EcoRV site is  destroyed
                      upon cloning).  Average insert size 1.5 kb, insert size
                      range 1-3 kb.  Library is normalized and enriched for
                      full-length  clones and was constructed by C. Gruber
                      (Invitrogen).  Research  Genetics tracking code 019. Note:
                      this is a NIH_MGC Library."
BASE COUNT      123 a    256 c    267 g     69 t
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 12;  Length 715;
  Best Local Similarity   58.9%;  Pred. No. 2.8e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        505 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 446

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        445 TCCTCCCCGGCGC 433


RESULT 10
BX342864/c
LOCUS       BX342864                1080 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX342864 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL007YC01 5-PRIME, mRNA sequence.
ACCESSION   BX342864
VERSION     BX342864.1  GI:30340126
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1515.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DL007AB01QP1&cluster=1515.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DL007AB01QP1.
FEATURES              Location/Qualifiers
     source           1. .1080
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DL007YC01"
                      /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                      /cell_line="RAMOS CELL LINE"
                      /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                      25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      131 a    402 c    386 g    132 t     29 others
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 13;  Length 1080;
  Best Local Similarity   58.9%;  Pred. No. 3e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
```

```
                    | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        541 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 482

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        481 TCCTCCCCGGCGC 469


RESULT 11
BX397064/c
LOCUS       BX397064                1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX397064 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI028YL10 5-PRIME, mRNA sequence.
ACCESSION   BX397064
VERSION     BX397064.1  GI:30617181
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1515.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI028DF05QP1&cluster=1515.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI028DF05QP1.
FEATURES              Location/Qualifiers
     source           1. .1201
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI028YL10"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      132 a    401 c    408 g    197 t     63 others
ORIGIN

  Query Match             39.7%;  Score 29.8;  DB 13;  Length 1201;
  Best Local Similarity   58.9%;  Pred. No. 3e+02;
  Matches   43;  Conservative    3;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          2 GACUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCC 61
              | | ||  | | |  || | | :| |   ||||| |||| ||||||| | | || |  ||
Db        334 GCCGCCTTCGCAGAGCCCCTCGTCGCCTTCCCAGAGCCCGGACAGAGGGGCGCCTCCGCC 275

Qy         62 UCCUCUGAGACGC 74
              :||:|   | |||
Db        274 TCCTCCCCGGCGC 262


RESULT 12
BI724568/c
LOCUS       BI724568                 407 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031072H08.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI724568
VERSION     BI724568.1  GI:15700263
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
```

```
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES              Location/Qualifiers
     source           1. .407
                      /organism="Chlamydomonas reinhardtii"
                      /mol_type="mRNA"
                      /strain="CC-1690 wild type mt+ 21gr"
                      /db_xref="taxon:3055"
                      /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                      ), Lambda Zap II"
                      /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                      XhoI; Stress condition II library, constructed by John
                      Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                      cells grown to mid-log phase in TAP (NH4+ - containing)
                      and shifted to TAP - NO3- (24hrs); H2 production
                      conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                      Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                      sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                      PolyA mRNA was purified from each sample, pooled and cDNA
                      synthesized. The cDNA was directionally cloned into lambda
                      Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                      sites. pBluescript II SK- plasmids were excised from the
                      lambda ZAP clones by superinfection with ExAssist
                      (Stratagene) phage. The library was normalized using
                      method 4 described in Bonaldo et al., (1996) Genome
                      Research 6: 791-806."
BASE COUNT      51 a    110 c    185 g     61 t
ORIGIN

  Query Match             39.5%;  Score 29.6;  DB 12;  Length 407;
  Best Local Similarity   65.0%;  Pred. No. 2.9e+02;
  Matches   39;  Conservative    2;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 CUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUC 63
              |: ||||| |||||  ||   || ||||  |||||||     || |:|  ||| ||| |
Db         73 CTGCAGCCCCGGACCCAGCCGGCCCAGCCACGCGCCCAGCGCCAGCGTCAGCACGCCCGC 14


RESULT 13
BM003088/c
LOCUS       BM003088                 516 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  1031108E01.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BM003088
VERSION     BM003088.1  GI:16437868
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished
COMMENT     Contact: Charles Hauser
```

DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES Location/Qualifiers
source 1. .516
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 58 a 143 c 237 g 78 t
ORIGIN

Query Match 39.5%; Score 29.6; DB 12; Length 516;
Best Local Similarity 65.0%; Pred. No. 3e+02;
Matches 39; Conservative 2; Mismatches 19; Indels 0; Gaps 0;

```
Qy          4 CUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUC 63
              |: |||||| |||||  ||   || ||||  ||||||||    || |:|  ||| ||| |
Db         76 CTGCAGCCCCGGACCCAGCCGGCCCAGCCACGCGCCCAGCGCCAGCGTCAGCACGCCCGC 17
```

RESULT 14
BZ639961
LOCUS BZ639961 563 bp DNA linear GSS 29-JAN-2003
DEFINITION OGCCP26TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0141F04, genomic survey sequence.
ACCESSION BZ639961
VERSION BZ639961.1 GI:28095052
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 563)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .563
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0141F04"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
BASE COUNT 152 a 132 c 132 g 147 t
ORIGIN

Query Match 39.2%; Score 29.4; DB 29; Length 563;
Best Local Similarity 56.3%; Pred. No. 3.4e+02;
Matches 40; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

```
Qy          5 UCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUCC 64
              :||| |||:|| |    | |:|   | ||||  ||| || |  |  | ||||||| | |:
Db        337 TCCAACCCTGGGCACTGGATCATAATCCCACGGCCAAGGCTTAACGCCCCCACGCTCTAT 396

Qy         65 UCUGAGACGCC 75
               |: | |||||
Db        397 CCTTATACGCC 407
```

RESULT 15
BF616342/c
LOCUS BF616342 867 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEc0006L10f Hordeum vulgare seedling shoot EST library HVcDNA0003 (Etiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA clone HVSMEc0006L10f, mRNA sequence.
ACCESSION BF616342
VERSION BF616342.2 GI:13107217
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 867)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling shoot cDNA library
JOURNAL Unpublished
COMMENT On Dec 18, 2000 this sequence version replaced gi:11880076.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 175
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 783.
FEATURES Location/Qualifiers
source 1. .867
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/db_xref="taxon:112509"
/clone="HVSMEc0006L10f"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare seedling shoot EST library HVcDNA0003 (Etiolated and unstressed)"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript

SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 153 a 123 c 408 g 183 t
ORIGIN

Query Match 39.2%; Score 29.4; DB 10; Length 867;
Best Local Similarity 61.9%; Pred. No. 3.7e+02;
Matches 39; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

```
Qy          4 CUCCAGCCCUGGACCGCGCAUCCCGAGCCCAGCGCCCAGACAGAGUGUCCCCACACCCUC 63
              | ||  |||  |||||  || ||||| ||| || | |  ||    :  |||| |||||:|
Db        832 CGCCGCCCCCCGACCGATCAACCCGACCCCTGCCCACCCACCACCTAACCCCCCACCCTC 773

Qy         64 CUC 66
              |:|
Db        772 CTC 770
```

Search completed: November 13, 2003, 13:52:38
Job time : 2055 secs